MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A424 529	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE RATE-DISTORTION FUNCTION ON CLASSES OF SOURCES DETERMINED BY SPECTRAL CAPACITIES		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER R-905;UILU-ENG 81-2236
7. AUTHOR(s) H. Vincent Poor		8. CONTRACT OR GRANT NUMBER(s) ECS79-16453;N00014-79-C-0424
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory University of Illinois at Urbana-Champaign Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS National Science Foundation; Joint Services Electronics Program		12. REPORT DATE April 1981
		13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Rate-distortion functions; spectral uncertainty; classes of sources;

Choquet capacities

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The quantity $\sup_{a \in \mathcal{Q}} R_a(D)$ is considered, where $\mathcal{Q}$ is a class of homogeneous n-parameter sources and $R_a(D)$ denotes the single-letter MSE rate-distortion function for the individual source a. In particular, the case in which the class $\mathcal{Q}$ is specified in terms of spectral information is treated for general class of spectral measures whose upper measures are capacities (in the sense of Choquet) alternating of order two. This type of class includes many common models for spectral uncertainty such as mixture models, spectral band models, and neighborhoods (over)

DD FORM 1473
1 JAN 73

20. (continued)

generated by Kolmogorov (total-variation) and Prohorov metrics. It is shown
that each such class contains a worst-case source whose rate-distortion
function achieves the supremum over the class for each value of distortion.
This source is characterized as having a spectral density that is a derivative
(in the sense of Huber and Strassen) of the upper spectral measure with respect
to Lebesgue measure on $[-\pi,\pi]^n$. Moreover it is shown that the spectral measure
of the worst-case source is closest, in a sense defined by directed divergence,
to Lebesgue measure (which corresponds to a memoryless source). Numerical
results are presented for the particular case in which the source spectral
measure is a mixture of a Gauss-Markov spectrum and an unknown contaminating
component.

FOREWORD

This report is a preprint of a paper with the same title which is

scheduled to appear in the January 1982, issue of the IEEE Transactions

on Information Theory (vol. IT-28). This work was presented at the 1981

IEEE International Symposium on Information Theory, Santa Monica, CA,

February 9-12, 1981.

THE RATE-DISTORTION FUNCTION ON CLASSES OF SOURCES
DETERMINED BY SPECTRAL CAPACITIES

by

H. Vincent Poor

I. INTRODUCTION

Suppose $\mathcal{Q}$ is a class of information sources. For a given distortion measure, consider the function

$$\sup_{a \in \mathcal{Q}} R_a(D); \quad D \geq 0 \tag{1}$$

where $R_a(D)$ is the rate-distortion function (corresponding to the given distortion measure) for the individual source a. The quantity in (1) is of interest in the study of coding schemes for classes of sources. For example, under compactness conditions on the class $\mathcal{Q}$ the quantity in (1) represents a rate-distortion function for $\mathcal{Q}$ (as defined by Sakrison [1]), and, even for noncompact classes, the quantity of (1) is a lower bound on this rate-distortion function. Further, if there is a universal code for $\mathcal{Q}$, then the quantity in (1) describes the worst-case rate (versus distortion) required to transmit, via the universal code, an arbitrarily chosen member of $\mathcal{Q}$.

In this paper we consider the function in (1) for classes of homogeneous discrete-parameter sources that are specified only in terms of the spectral properties of their elements. We consider the particular version of (1) corresponding to the single-letter mean-square-error (MSE) distortion measure. Note that for the one-parameter case, which corresponds to $\mathcal{Q}$ being a class of covariance-stationary discrete-time sources specified only in spectral terms, the MSE version of the function in (1) is unchanged if $\mathcal{Q}$ is replaced by its subset consisting only of Gaussian sources (Berger [2, p. 154]). Thus, in this context, it is reasonable to restrict attention to the consideration of classes of Gaussian sources. Such a class $\mathcal{Q}$ can be specified completely by defining a class $\mathcal{M}$ of spectral measures, and the determination of the quantity of (1) for two specific source classes of

this type has been considered previously by Sakrison [3]. Here, for a general type of such classes, we demonstrate the existence of a member spectrum whose rate-distortion function achieves (1) for each value of $D \geq 0$. In particular, we consider classes $\mathcal{M}$ whose upper measures are Choquet alternating capacities of order 2 (Choquet [4]). Such classes occupy a central position in generalizations of the theories of hypothesis testing (Huber and Strassen [5]) and of stationary linear smoothing (Poor [6]), and include many important classes such as contaminated mixtures, variation neighborhoods, and Prohorov neighborhoods. We show here that the spectrum achieving the supremum in (1) for such a class is given by the derivative (in the sense of Huber and Strassen [5]) with respect to Lebesgue measure of the upper measure of $\mathcal{M}$ and corresponds to the element of $\mathcal{M}$ that is closest to Lebesgue measure in a sense defined by directed divergence. Since Lebesgue measure represents a white spectrum (corresponding to a memoryless source), the maximizing spectrum is thus the member of $\mathcal{M}$ that is "most memoryless", a result that certainly agrees with the intuitive meaning of the MSE rate-distortion function.

Section II contains a more complete specification of the spectral classes to be considered and gives properties of these classes that are relevant to the determination of the quantity $\sup_{a \in \mathcal{Q}} R_a(D)$. A number of examples are also presented that demonstrate the generality of this type of class. Section III contains the main analytical results concerning the quantity in (1) for such classes. In particular the existence of a maximizing spectrum is demonstrated, and the characterization of this spectrum as the element of $\mathcal{M}$ closest to Lebesgue measure is established. Section IV considers in detail the specific case in which the class $\mathcal{M}$ consists of all spectra that are a convex mixture of a discrete-time wide-sense Markov spectrum and an unknown "contaminating" spectrum, and the results of

Section III are illustrated directly for this case. The extension of the results to continuous-parameter cases is discussed in Section V.

II. CHOQUET CAPACITIES, HUBER-STRASSEN DERIVATIVES, AND SPECTRAL UNCERTAINTY

Throughout this paper n is a fixed positive integer and $\underline{X} = \{X_t ; t \in Z^n\}$ is an n-parameter homogeneous Gaussian source (Z denotes the set of all integers). By way of Bochner's theorem (Wong [7, p. 245]) a class of sources of this type can be specified by defining a class $\mathcal{M}$ of spectral measures on $(\Omega, \mathcal{B})$ where Ω denotes the n-dimensional rectangle $[-\pi, \pi]^n$ and $\mathcal{B}$ denotes the Borel σ-algebra on Ω. In this paper we will restrict our attention primarily to classes $\mathcal{M}$ satisfying the power constraint $m(\Omega) = (2\pi)^{-n} P$ for all $m \in \mathcal{M}$ where P is a fixed positive number. It is straightforward to relax this constraint; this restriction, however, allows us to consider the effects of spectral shape on the rate-distortion function in more detail.

The _upper measure_ v of a class $\mathcal{M}$ with power constraint is the set function on $\mathcal{B}$ defined by

$$v(B) = \sup_{m \in \mathcal{M}} m(B); \quad B \in \mathcal{B}. \tag{2}$$

Note that v has the following properties: (i) $v(\phi) = 0$ and $v(\Omega) = (2\pi)^{-n} P$, (ii) $A \subset B$ implies $v(A) \leq v(B)$, and (iii) $B_n \uparrow B$ implies $v(B_n) \uparrow v(B)$, where all sets are assumed to be in $\mathcal{B}$ and ϕ denotes the null set. If $\mathcal{M}$ is weakly compact, then v has the additional property (Huber and Strassen [5, p. 252]): (iv) $F_n \downarrow F$ with F_n closed for all n implies $v(F_n) \downarrow v(F)$. A set function having properties (i) through (iv) is a _capacity_ on $\mathcal{B}$ in the sense of Choquet [4]. In this paper we consider classes whose upper measures satisfy (i)-(iv) and the additional property: (v) $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$ for all $A, B \in \mathcal{B}$. A capacity satisfying this latter property is said to be alternating of order 2 and is termed a _2-alternating capacity_. Note that a finite measure is an example of a 2-alternating capacity.

If the upper measure v of a class $\mathcal{M}$ is a 2-alternating capacity then $\mathcal{M}$ coincides with the class $\mathcal{M}_v$ defined by

$$\mathcal{M}_v = \{m \in M \mid m(B) \leq v(B); \ B \in \mathcal{B}, \text{ and } m(\Omega) = v(\Omega)\} \tag{3}$$

where M denotes the class of all finite measures on $(\Omega, \mathcal{B})$ (Huber and Strassen [5, Lemma 2.5]). Classes of the form of (3) include many of the traditional models for spectral uncertainty, and a number of useful examples are discussed below. It is interesting to note that all such classes are weakly compact [5, Lemma 2.2] and that, if v is a measure, then $\mathcal{M}_v = \{v\}$. Thus any results obtained for $\mathcal{M}_v$ also apply to a single measure; in fact the classes $\mathcal{M}_v$ have been useful in generalizing hypothesis testing and Wiener filtering results which hold for single measures (such as the Neyman-Pearson Lemma) to classes of measures [5,6].

The properties of classes of the form of (3) have been studied by Huber and Strassen [5]. The properties of such classes relevant to the present problem can be summarized in the following two lemmas (here, and elsewhere in this paper, λ denotes Lebesgue measure on $(\Omega, \mathcal{B})$):

Lemma 1: Suppose v is a 2-alternating capacity on $(\Omega, \mathcal{B})$. There exists a Lebesgue-measurable function $\pi_v : \Omega \to [0, \infty]$ such that[1] $w_\theta(\{\pi_v > \theta\}) = \inf_{B \in \mathcal{B}} w_\theta(B)$ for each $\theta \geq 0$ where w_θ is the set function on $\mathcal{B}$ defined by

$$w_\theta(B) = v(B^c) + \theta\lambda(B), \quad B \in \mathcal{B}. \tag{4}$$

Furthermore, π_v is unique a.e. $[\lambda]$.

---

[1] For compactness of notation we will write $\{f > \theta\}$ to denote $\{\omega \in \Omega \mid f(\omega) > \theta\}$.

Proof: Since λ is a finite measure and hence a 2-alternating capacity on $(\Omega, \mathcal{B})$, the existence of π_v follows from Lemmas 3.1 and 3.2 of [5]. The uniqueness of π_v follows from Theorem 5.1 of [5].

Lemma 2: Suppose v and π_v are as in Lemma 1 and $B_\theta \triangleq \{\pi_v > \theta\}$, $\forall\, \theta \geq 0$. Then there exists a measure $q \in \mathcal{M}_v$ such that $\pi_v = dq/d\lambda$ and $q(B_\theta^c) = v(B_\theta^c)$ for all $\theta \geq 0$. Here $dq/d\lambda$ denotes the generalized Radon-Nikodym derivative of q with respect to λ; that is, $dq/d\lambda$ may be infinite on a set of λ measure zero.

Proof: The existence of such a q follows from Theorem 4.1 of [5] and from the construction in the proof of this theorem (see also Huber and Strassen [8]).

Note that, if v is a finite measure, then the function π_v is the (generalized) Radon-Nikodym derivative of v with respect to λ. Partly for this reason Huber and Strassen termed the function π_v the Radon-Nikodym derivative of v with respect to λ. However, to distinguish π_v from an ordinary Radon-Nikodym derivative, we will term π_v the <u>Huber-Strassen</u> <u>derivative</u> of v with respect to λ. Lemmas 1 and 2 are particular cases of more general results for differentiating one capacity with respect to another [5]. In this more general context, the Huber-Strassen derivative is the basis for minimax hypothesis testing between two classes of probability measures of the form $\mathcal{M}_v$ [5] and for minimax linear smoothing of a signal with spectral measure in a class of the form $\mathcal{M}_v$ observed in additive noise with spectral measure in a class of the form $\mathcal{M}_v$ [6].

Lemmas 1 and 2 give the basic properties needed to consider the rate-distortion function over classes of Gaussian sources determined by spectral classes of the form of (3). We conclude this section by giving several useful examples of such classes that illustrate the generality of this

model. In Examples 1 and 2, m_0 is a fixed finite spectral measure on $(\Omega, \mathcal{B})$ and ϵ is a fixed number in $[0,1]$. In all cases $v_i(\phi)$ is defined to be zero.

__Example 1 (ϵ-mixtures)__: The set function $v_1(B) = (1 - \epsilon)m_0(B) + \epsilon m_0(\Omega)$, $B \in \mathcal{B}$, $B \neq \phi$ is a 2-alternating capacity[2] and $\mathcal{M}_{v_1}$ is given by

$$\mathcal{M}_{v_1} = \{m \in M \mid m = (1 - \epsilon)m_0 + \epsilon h \text{ for some } h \in M \text{ with } h(\Omega) = m_0(\Omega)\}. \qquad (5)$$

Thus m_0 can be thought of as a nominal spectral model and ϵ as a degree of uncertainty placed on the model. The model of (5) is one of the earliest models used for uncertainty in robust hypothesis testing and signal-detection studies (Huber [9], Martin and Schwartz [10]) and was noted in [5] to be of the form of (3). The Huber-Strassen derivative of v_1 with respect to Lebesgue measure on Ω is found by noting that the set function w_θ of (4) is given for this case by

$$w_\theta(B) = \begin{cases} (1 - \epsilon)m_0(B^c) + \epsilon m_0(\Omega) + \theta\lambda(B); & B \neq \Omega \\ \\ \theta\lambda(\Omega) = \theta(2\pi)^n & ; \quad B = \Omega \end{cases} \qquad (6)$$

which is minimized over $\mathcal{B}$ by the set

---

[2] Note that this set function (and the one defined in Example 2) is discontinuous from above at the null set ϕ. However, for $\Omega = [-\pi, \pi]^n$, this discontinuity does not violate the property that v must be continuous from above on closed sets (i.e., Property (iv)) since, in a compact separable metric space, there is no sequence of closed sets converging down to the null set (Dunford and Schwartz [11, pp. 30-31]).

$$B_\theta = \begin{cases} \{\pi_0 > \theta/(1-\epsilon)\}; & \text{if } w_\theta(\{\pi_0 > \theta/(1-\epsilon)\}) \le \theta_{1})^n \\[2em] \Omega & ; \text{ otherwise} \end{cases} \qquad (7)$$

where $\pi_0 = dm_0/d\lambda$. Noting that we can define $\pi_v(\omega) = \inf\{\theta \ge 0 | \omega \notin B_\theta\}$, we have that

$$\pi_{v_1}(\omega) = \max\{c', (1-\epsilon)\pi_0(\omega)\}, \quad \omega \in \Omega, \qquad (8)$$

where $c' = \sup\{\theta \ge 0 | w_\theta(\{\pi_0 > \theta/(1-\epsilon)\}) \le \theta(2\pi)^n\}$. A specific example of this class will be considered in Section IV below.

Example 2 (variation neighborhoods): The set function $v_2(B) = \min\{m_0(B) + \epsilon m_0(\Omega), m_0(\Omega)\}$, $B \in \mathcal{B}$, $B \ne \phi$, is a 2-alternating capacity with $\mathcal{M}_{v_2}$ given by

$$\mathcal{M}_{v_2} = \{m \in M | \rho(m, m_0) \le \epsilon m_0(\Omega), \text{ and } m(\Omega) = m_0(\Omega)\} \qquad (9)$$

where ρ is the variational distance (or Kolmogorov metric) defined on M by $\rho(\mu, \nu) = \sup_{B \in \mathcal{B}}|\mu(B) - \nu(B)|$. Classes of the form of (9) have been considered previously in the contexts of robust hypothesis testing (Huber [9]) and robust Wiener filtering (Poor [12]). Note that this class is also a model for a degree ϵ of uncertainty in a nominal spectral model m_0, although the deviation allowed in (9) is somewhat different from that allowed in (5). Note further that (9) with degree of uncertainty $\epsilon/2$ contains (5) with degree of uncertainty ϵ. The derivative of v_2 with respect to Lebesgue measure on Ω can be found from general results in [13] and [14] and is given by

$$\pi_{v_2}(\omega) = \max\{c', \min\{c'', \pi_0(\omega)\}\}; \quad \omega \in \Omega, \tag{10}$$

where $\pi_0 = dm_0/d\lambda$, $c'' = \inf\{\theta \geq 0 \,|\, m_0(\{\pi_0 > \theta\}) - \epsilon m_0(\Omega) < \theta\lambda(\{\pi_0 > \theta\})\}$, and $c' = \sup\{\theta \geq 0 \,|\, \theta\lambda(\{\pi_0 \leq \theta\}) \leq m_0(\{\pi_0 \leq \theta\}) + \epsilon m_0(\Omega)\}$.

<u>Example 3 (band models)</u>: Suppose $P > 0$ is the power in $\underline{X}$ and that m_L and m_U are two finite spectral measures on $(\Omega, \mathcal{B})$ with $m_L(\Omega) < (2\pi)^n P < m_U(\Omega)$. Define two 2-alternating capacities v_L and v_U by $v_L(B) = m_L(B) + ((2\pi)^n P - m_L(\Omega))$, $B \in \mathcal{B}$, $B \neq \phi$, and $v_U(B) = \min\{m_U(B), (2\pi)^n P\}$, $B \in \mathcal{B}$, $B \neq \phi$. Then the set function $v_3(B) = \min\{v_L(B), v_U(B)\}$ is a 2-alternating capacity on $(\Omega, \mathcal{B})$ and $\mathcal{M}_{v_3}$ is given by

$$\mathcal{M}_{v_3} = \{m \in M \,|\, m_L \leq m \leq m_U \text{ and } m(\Omega) = (2\pi)^n P\}. \tag{11}$$

Thus, $\mathcal{M}_{v_3}$ for this case is the band of spectral measures lying between the spectral measures m_L and m_U. This class is known as the <u>band model</u> and has been utilized previously in problems of robust signal detection (Kuznetsov [15], Kassam [16]) and robust Wiener filtering (Kassam and Lim [17]). This class was shown to be of the type (3) by Vastola and Poor [18]. Note that, with $m_L = (1 - \epsilon)m_0$ and with $m_U \rightarrow \infty$, the class of (11) gives the ϵ-mixture class of (5). The derivative of v_3 with respect to Lebesgue measure on Ω can be determined from a result in [18] and involves both derivatives $dm_U/d\lambda$ and $dm_L/d\lambda$.

Other examples of classes of the form of (3) include Prohorov neighborhoods of a nominal spectral measure m_0 (see Strassen [19, p. 438]) and generalizations of the mixture model of Example 1 generated by mixed capacities of the form $v(B) = \int_{\mathcal{G}} v_g(B)\mu(dg)$, where μ is a measure on $\mathcal{G}$ (see Strassen [19, Theorem 4]). A class of spectral measures that does not fit the model of (3) is Sakrison's Model (b) (see [3, p. 11]) which

essentially consists of those measures placing exact amounts of spectral measure on members of a set of intervals covering the interval $[-\pi, \pi]$. This type of model is not of the type in (3) because such classes of measures are not weakly closed (Vastola and Poor [18]).

III. THE RATE-DISTORTION FUNCTION ON SPECTRAL CLASSES GENERATED BY CAPACITIES

The MSE rate-distortion function for a homogeneous Gaussian source $\underline{X}$ with spectral measure m depends only on the part of m that is absolutely continuous with respect to Lebesgue measure. This function, denoted by $R_m(D)$, is given by the parametric relationship

$$R_m(D_\theta(m)) = (4\pi)^{-n} \int_\Omega \max\{0, \log(\pi_m(\omega)/\theta)\}\lambda(d\omega) \tag{12}$$

and

$$D_\theta(m) = (2\pi)^{-n} \int_\Omega \min\{\theta, \pi_m(\omega)\}\lambda(d\omega) \tag{13}$$

where $\pi_m = dm/d\lambda$ is the generalized Radon-Nikodym derivative of m with respect to λ, and θ is a parameter ranging over the interval $(0, \Delta_m]$ with $\Delta_m =$ ess sup$_{\omega \in \Omega}$ $\pi_m(\omega)$. Note that $0 \le \Delta_m < \infty$, since Ω is compact. The expression of (12) and (13) is given by Berger [2, Theorem 4.5.3] for the case n = 1 and follows from Hayes, Habibi, and Wintz [20] for n $\ge$ 2.

It follows from (12) and (13) that the support of $R_m(D)$ is the interval $(0, D_m)$ and that $R_m(D_m) = 0$, where $D_m = (2\pi)^{-n} \int_\Omega \pi_m d\lambda$. With respect to this latter quantity we have the following result:

Lemma 3: Suppose v is a 2-alternating capacity on $(\Omega, \mathcal{B})$ with Huber-Strassen derivative π_v with respect to λ. Then $\int_\Omega \pi_v d\lambda \ge \int_\Omega \pi_m d\lambda$ for all $m \in \mathcal{M}_v$.

<u>Proof</u>: Suppose $m \in \mathcal{M}_v$. For each $\vartheta \geq 0$, define sets $B_\vartheta = \{\pi_v > \theta\}$ and $A_\vartheta = \{\pi_m > \theta\}$. By Lemma 1 we have

$$m(A_\vartheta^c) + \theta \lambda(A_\vartheta) = \inf_{B \in \mathcal{B}} [m(B^c) + \theta \lambda(B)] \leq m(B_\vartheta^c) + \theta \lambda(B_\vartheta)$$

$$\leq v(B_\vartheta^c) + \theta \lambda(B_\vartheta) \tag{14}$$

for each $\vartheta \geq 0$, where the second inequality follows from the fact that $m \leq v$. Lemma 2 states that there is a measure q such that π_v is a version of $dq/d\lambda$ and $v(B_\vartheta^c) = q(B_\vartheta^c)$. Thus, (14) implies

$$\int_{A_\vartheta^c} \pi_m d\lambda \leq \int_{B_\vartheta^c} \pi_v d\lambda + \theta [\lambda(B_\vartheta) - \lambda(A_\vartheta)]$$

for each $\theta < \infty$, which in turn implies

$$\int_{A_\vartheta^c} \pi_m d\lambda \leq \int_{B_\vartheta^c} \pi_v d\lambda + \int_{B_\vartheta} \pi_v d\lambda + \int_{A_\vartheta} \pi_m d\lambda \tag{15}$$

for each $\theta < \infty$. Taking the limit as $\vartheta \to \infty$ in (15) and defining $B_\infty = \bigcap_{\theta > 0} B_\vartheta$ and $A_\infty = \bigcap_{\vartheta > 0} A_\vartheta$, we have

$$\int_{A_\infty^c} \pi_m d\lambda \leq \int_{B_\infty^c} \pi_v d\lambda + \int_{B_\infty} \pi_v d\lambda + \int_{A_\infty} \pi_m d\lambda . \tag{16}$$

Since A_∞ and B_∞ both have λ measure zero, (16) is equivalent to $\int_\Omega \pi_m d\lambda \leq \int_\Omega \pi_v d\lambda$ which was to be shown.

We see from Lemma 3 that no $m \in \mathcal{M}_v$ has more nonsingular power than the member q singled out by Lemma 2. What we will show in the following paragraphs is that $R_q(D)$ achieves $\sup_{m \in \mathcal{M}_v} R_m(D)$, and hence that the spectral density defined by π_v also achieves this supremum. To do this, we first

need the following result which seems intuitively obvious from the derivation of (12) and (13) and from analogous results for memoryless sources (see Berger [2, Theorem 4.2.1 and pp. 110-111]) but for which we could find no previously published proof.

Lemma 4: Suppose m is a spectral measure on $(\Omega,\mathcal{B})$ with MSE rate-distortion function $R_m(D)$. Then $R_m(D)$ has a right-hand derivative $R_m'(D)$ for all $D \in (0,D_m)$ where $D_m = (2\pi)^{-n} \int_\Omega \pi_m d\lambda$, and this derivative is given by $R_m'(D) = -2^{-n}\theta_D^{-1}$ where θ_D is determined uniquely by the equation

$$D = (2\pi)^{-n} \int_\Omega \min\{\theta_D, \pi_m(\omega)\}\lambda(d\omega). \tag{17}$$

Proof: The existence of the right-hand derivative of $R_m(D)$ follows from the fact that $R_m(D)$ is convex on $(0,D_m)$ (see Berger [2, p. 270] and Royden [21, p. 109]). The relationship $D_\theta(m) = (2\pi)^{-n}[m(\{\pi_m \leq \theta\}) + \theta\lambda(\{\pi_m > \theta\})]$ defines a continuous, strictly increasing mapping from $(0, \Delta_m)$ onto $(0, D_m)$. This follows by noting that, for $0 < \theta < \theta' < \Delta_m$, we have

$$(\theta'-\theta)\lambda(\{\pi_m > \theta\}) \geq (\theta'-\theta)\lambda(\{\pi_m > \theta\}) - \int_{\{\theta < \pi_m \leq \theta'\}} (\theta'-\pi_m)d\lambda =$$
$$= (2\pi)^n[D_{\theta'}(m) - D_\theta(m)] = (\theta'-\theta)\lambda(\{\pi_m > \theta'\}) + \int_{\{\theta < \pi_m \leq \theta'\}} (\pi_m - \theta)d\lambda$$
$$\geq (\theta'-\theta)\lambda(\{\pi_m > \theta'\}) > 0, \tag{18}$$

where the final inequality follows from the fact that $\theta' < \Delta_m = \text{ess sup}_{\omega \in \Omega} \pi_m(\omega)$. Thus, a unique θ_D is determined by (17) for each $D \in (0,D_m)$, and the right-hand derivative of $R_m(D)$ is given by the expression

$$R_m'(D) = \lim_{\theta \downarrow \theta_D} [(R_m(D_\theta(m)) - R_m(D))/(D_\theta(m) - D)]. \tag{19}$$

Equation (12) implies that the numerator in (19) can be written as

$$R_m(D_\theta(m)) - R_m(D) = (4\pi)^{-n}[\log(\theta_D/\theta)\lambda(\{\pi_m > \theta\}) + \int_{\{\theta_D < \pi_m \leq \theta\}} \log(\theta_D/\pi_m)d\lambda]$$

$$= (4\pi)^{-n}[\log(\theta_D/\theta)\lambda(\{\pi_m > \theta\}) - \theta_D^{-1}\int_{\{\theta_D < \pi_m \leq \theta\}}(\pi_m - \theta)d\lambda + 0[(\theta - \theta_D)^2]].$$

$$(20)$$

Similarly the equality in the middle of (18) gives

$$D_\theta(m) - D = (2\pi)^{-n}[(\theta - \theta_D)\lambda(\{\pi_m > \theta\}) + \int_{\{\theta_D < \pi_m \leq \theta\}}(\pi_m - \theta_D)d\lambda]. \qquad (21)$$

We thus have

$$|(R_m(D_\theta(m)) - R_m(D))/(D_\theta(m) - D) + 2^{-n}\theta_D^{-1}| =$$

$$(4\pi)^{-n}|(\log(\theta_D/\theta) + \theta_D^{-1})\lambda(\{\pi_m > \theta\}) + 0((\theta - \theta_D)^2)|/(D_\theta(m) - D)$$

$$\leq 2^{-n}|(\log(\theta_D/\theta) + \theta_D^{-1})/(\theta - \theta_D) + 0(\theta - \theta_D)/\lambda(\{\pi_m > \theta\})|. \qquad (22)$$

Since $\lim_{\theta \to \theta_D}|(\log(\theta_D/\theta) + \theta_D^{-1})/(\theta - \theta_D)| = 0$ and $\lambda(\{\pi_m > \theta_D\}) > 0$, the

limit as θ approaches θ_D of the right-hand side of (22) is zero.

Thus $R_m'(D) = -2^{-n}\theta_D^{-1}$, which was to be shown.

We may now use Lemma 4 to prove the main analytical result of this

paper. In particular, we have the following:

Theorem 1: Suppose v is a 2-alternating capacity on (Ω, β). Define $P = (2\pi)^{-n}v(\Omega)$ and $P_v = (2\pi)^{-n}\int_\Omega \pi_v d\lambda$, where $\pi_v = dv/d\lambda$. Then $\sup_{m \in \mathcal{M}_v} R_m(D)$ is

15

defined parametrically on $(0,P_v)$ by the equations

$$\sup_{m \in \mathcal{M}_v} R_m(D_\theta) = (4\pi)^{-n} \int_\Omega \max\{0, \log(\pi_v(\omega)/\theta)\}\lambda(d\omega) \qquad (23)$$

and

$$D_\theta = (2\pi)^{-n} \int_\Omega \min\{\theta, \pi_v(\omega)\}\lambda(d\omega), \qquad (24)$$

and $\sup_{m \in \mathcal{M}_v} R_m(D) = 0$ for $D \in [P_v, P]$. Moreover, there exists a measure $q \in \mathcal{M}_v$ such that $\sup_{m \in \mathcal{M}_v} R_m(D) = R_q(D)$ and $dq/d\lambda = \pi_v$ a.e. $[\lambda]$.

Proof: Suppose $m \in \mathcal{M}_v$; then $R_m(D)$ is absolutely continuous on any interval $[D,P]$ with $D > 0$ since it is convex on $(0,P]$ (Royden [21, p. 109]). Thus, for any $D \in (0,P)$,

$$R_m(D) = R_m(P) - \int_D^P R_m'(x)dx = -\int_D^P R_m'(x)dx, \qquad (25)$$

where $R_m'(x)$ is obtained from Lemma 4 for $x \in (0, P_m)$ and $R_m'(x) = 0$ for $x \geq P_m$. Equation (25) is equivalent to

$$R_m(D) = \begin{cases} 2^{-n} \int_D^{P_m} \theta_x^{-1}(m)dx; & D \in (0, P_m) \\ 0 & ; & D \in [P_m, P] \end{cases} \qquad (26)$$

where $\theta_x(m)$ is the unique value of θ solving $x = (2\pi)^{-n}[m(\{\pi_m \leq \theta\}) + \theta\lambda(\{\pi_m > \theta\})]$. Note that the existence and uniqueness of $\theta_x(m)$ follows from the development in the proof to Lemma 4. Suppose $q \in \mathcal{M}_v$ is such that $dq/d\lambda = \pi_v$ a.e. $[\lambda]$ and $q(\{\pi_v \leq \theta\}) = v(\{\pi_v \leq \theta\})$ for all $\theta \geq 0$ (the existence of such a q follows from Lemma 2). Then

$$m(\{\pi_m \le \theta\}) + \theta\lambda(\{\pi_m > \theta\}) = \inf_{B \in \mathcal{B}}[m(B^c) + \theta\lambda(B)] \le \inf_{B \in \mathcal{B}}[v(B^c) + \theta\lambda(B)]$$

$$= v(\{\pi_v \le \theta\}) + \theta\lambda(\{\pi_v > \theta\}) = q(\{\pi_q \le \theta\}) + \theta\lambda(\{\pi_q > \theta\}).$$

$$(27)$$

Thus, since $[m(\{\pi_m \le \theta\}) + \theta\lambda(\{\pi_m > \theta\})]$ is strictly increasing in θ on $(0, \Delta_m)$, we must have $\theta_x(q) \le \theta_x(m)$ for all $x \in (0, P_m)$. Therefore, we have $\theta_x^{-1}(q) \ge \theta_x^{-1}(m)$ for all $x \in (0, P_m)$ and, since Lemma 3 implies $P_q \ge P_m$, (26) gives $R_m(D) \le R_q(D)$ for all $D \in (0, P)$. To complete the proof we need only note that the quantity defined by (23) and (24) is $R_q(D)$ on $(0, P_q)$ and that $P_q = P_v$.

Theorem 1 implies that the quantity $\sup_{m \in \mathcal{M}_v} R_m(D)$ is the rate-distortion curve corresponding to the spectral density π_v, which is the Huber-Strassen derivative of v with respect to Lebesgue measure on $(\Omega, \mathcal{B})$. This theorem also states that there is a $q \in \mathcal{M}_v$ that has distortion rate $\sup_{m \in \mathcal{M}_v} R_m(D)$ for each $D \in (0, P]$. Thus, for classes of the form $\mathcal{M}$, the spectral density π_v represents a worst-case or least-favorable spectrum in terms of MSE distortion rate, and the problem of finding $\sup_{m \in \mathcal{M}_v} R_m(D)$ is solved once π_v is determined. For the examples given in Section II, the solution is thus obtained. In general, we have the following theorem which characterizes the spectral measures $q \in \mathcal{M}_v$ (with $dq/d\lambda = \pi_v$ a.e. $[\lambda]$) that are singled out by Lemma 2.

<u>Theorem 2</u>: For each measure $m \in \mathcal{M}_v$, define m' to be the part of m that is absolutely continuous with respect to λ (via the Lebesgue decomposition). A measure $q \in \mathcal{M}_v$ satisfies the conclusion of Lemma 2 if and only if, q minimizes the quantity

$$J(m) = \int_{\Omega} \log[d(\lambda + m')/d\lambda] d(\lambda + m') \qquad (27)$$

over $\mathcal{M}_v$.

Proof: Suppose q satisfies the conclusion of Lemma 2, i.e., $dq/d\lambda = \pi_v$ a.e. $[\lambda]$ and $q(\{\pi_v \le \theta\}) = v(\{\pi_v \le \theta\})$ for all $\theta \ge 0$. Noting that $\mathcal{M}_v$ is convex, define, for each $\gamma \in [0,1]$ and $m \in \mathcal{M}_v$, the measure $m_\gamma = (1-\gamma)q + \gamma m$ and the function $\pi_\gamma = d(m'_\gamma + \lambda)/d\lambda$. Since $x \log(x)$ is convex for $x \in (0,\infty)$ and since $m'_\gamma = [(1-\gamma)q' + \gamma m']$ we have that $J(m)$ is convex on $\mathcal{M}_v$. Thus a sufficient condition for q to minimize $J(m)$ over $\mathcal{M}_v$ is that $\delta J(m_\gamma)/\delta\gamma|_{\gamma=0}$ be nonnegative for every $m \in \mathcal{M}_v$ satisfying $J(m) < \infty$. Since $\log[\pi_\gamma(\omega)]\pi_\gamma(\omega)$ is convex in γ on $[0,1]$ for each $\omega \in \Omega$, we have

$$(\pi_1 - \pi_0)(1 + \log(\pi_0)) \le \gamma^{-1}(\pi_\gamma \log(\pi_\gamma) - \pi_0 \log(\pi_0))$$

$$\le \pi_1 \log(\pi_1) - \pi_0 \log(\pi_0), \quad \text{a.e. } [\lambda]. \qquad (28)$$

The left-most quantity of (28) is bounded below a.e. $[\lambda]$ and the right-most quantity is integrable with respect to λ. Thus since the term $\gamma^{-1}(\pi_\gamma \log(\pi_\gamma) - \pi_0 \log(\pi_0))$ converges monotonically to $(\pi_1 - \pi_0)(1 + \log(\pi_0))$ as $\gamma \downarrow 0$, the Monotone Convergence Theorem (Royden [21, p. 84]) implies that

$$\delta J(m_\gamma)/\delta\gamma|_{\gamma=0} = \int_{\Omega} \delta[\pi_\gamma \log(\pi_\gamma)]/\delta\gamma|_{\gamma=0} d\lambda$$

$$= \int_{\Omega} (\pi_1 - \pi_0)(1 + \log(\pi_0)) d\lambda$$

$$= \int_{\Omega} (1 + \log(1 + \pi_v))dm' - \int_{\Omega} (1 + \log(1 + \pi_v))dq'. \tag{29}$$

Since $m'(\{\pi_v \le \theta\}) \le m(\{\pi_v \le \theta\}) \le v(\{\pi_v \le \theta\}) = q(\{\pi_v \le \theta\}) = q'(\{\pi_v \le \theta\})$

for all $\theta \ge 0$, we see that π_v is stochastically smaller under q' than

under m'. Thus, since $(1 + \log(1 + \pi_v))$ is increasing in π_v, the quantity

in (29) is nonnegative, which is sufficient for q to minimize J(m).

We now see that if q is as in Lemma 2 it minimizes J(m) over $\mathcal{M}_v$.

Suppose $p \in \mathcal{M}_v$ also minimizes J(m) over $\mathcal{M}_v$. Then, since J(m) is

convex on $\mathcal{M}_v$, $J((1-\gamma)q + \gamma p)$ must be constant for $\gamma \in [0,1]$. This implies,

for instance, that $\delta^2 J((1-\gamma)q + \gamma p)/\delta\gamma^2|_{\gamma=0} = 0$. Applying analysis

similar to that yielding (29), we have

$$\delta^2 J((1-\gamma)q + \gamma p)/\delta\gamma^2|_{\gamma=0} = \int_{\Omega} [(\pi_p - \pi_v)^2/\pi_0]d\lambda = 0 , \tag{30}$$

where $\pi_p = dp/d\lambda$. Equation (30) implies that $\pi_q = \pi_v$ a.e. $[\lambda]$ and, hence,

that $p' = q' \mod[\lambda]$. Thus p also satisfies the conclusion of Lemma 2.

This completes the proof.

Note that the quantity J(m) of (27) is a measure of the distance or

divergence of the measure $(\lambda + m')$ from λ (see, for example, Kullback [22]).

Thus, we see that the least-favorable spectral measure q is that which is

closest to Lebesgue measure in this sense. Since Lebesgue measure on Ω

represents white noise (which corresponds to a memoryless source) we see

that q is, in a sense, the "most memoryless" element of $\mathcal{M}_v$. Note also that

Lemma 3 implies that q has the maximum possible nonsingular power of any

measure in $\mathcal{M}_v$. These phenomena agree well with the intuitive meaning of

the rate-distortion function since, in general, a memoryless source requires

the highest rate for a given degree of distortion when no other constraints

are placed on the source spectrum.

IV. EXAMPLE — MIXTURE CONTAMINATED GAUSS-MARKOV SOURCES

In this section we use a specific example of the mixture model of Example 1 in Section II to illustrate our results. In particular we consider a one-parameter source ($n = 1$) whose spectral measure m has a nominal first-order Markov spectral density with degree of mixture uncertainty ϵ. That is, we have $m = (1-\epsilon)m_0 + \epsilon h$ where h is unknown, $\epsilon \in [0,1]$, and where m_0 is given by

$$m_0(A) = P(1-r^2) \int_A (1-2r\cos(\omega) + r^2)^{-1} \lambda(d\omega) \, , \, A \in \mathcal{B} \, , \qquad (31)$$

with $|r| \leq 1$. Recall that $P > 0$ is the source power. This nominal source corresponds to a Gaussian source with autocorrelation function $E(X_i X_{i+k}) = P \cdot r^{|k|}$; $k \in Z$. We will assume in the following that $P = 1$ and $r > 0$.

The rate-distortion function corresponding to the nominal source m_0 is discussed in [2, pp. 113-115]. This function is given by

$$R_{m_0}(D) = \frac{1}{2} \log \left[(1-r^2)/D \right] \qquad (32)$$

for $0 < D \leq (1-r)/(1+r)$, and parametrically by

$$D_\theta(m_0) = 1 + \frac{\theta x_\theta}{\pi} - \frac{2}{\pi} \tan^{-1}(\frac{1+r}{1-r}\tan(\frac{x_\theta}{2})) \qquad (33)$$

and

$$R_{m_0}(D_\theta(m_0)) = \frac{x_\theta}{2\pi} \log (\frac{1-r^2}{\theta}) - \frac{1}{2\pi}[C\ell_2(2x_\theta + 2y_\theta)$$
$$- C\ell_2(2x_\theta) - C\ell_2(2y_\theta) - 2y_\theta \log(r)] \qquad (34)$$

for $(1-r)/(1+r) \leq D \leq 1$, which corresponds to $(1-r)/(1+r) \leq \theta \leq (1+r)/(1-r)$. Here

$$x_\theta = \cos^{-1}\left[\frac{r^2-1+\theta(1+r^2)}{1-\theta}\right] , \quad 0 \leq x_\theta \leq \pi, \tag{35}$$

$$y_\theta = \tan^{-1}\left[\frac{r \sin(x_\theta)}{1 - r\cos(x_\theta)}\right] ; \quad 0 \leq y_\theta \leq \pi/2, \tag{36}$$

and $C\ell_2$ denotes Clausen's integral defined by

$$C\ell_2(x) = -\int_0^x \log(2\sin(t/2))dt. \tag{37}$$

(Note: There apparently are some minor errors in Eq. (4.5.35) of [2] which should correspond to (34); however, (34) follows directly from (12) and Eq. (39) on p. 272 of Lewin [23].)

Equation (8) specifies the worst-case spectral density (i.e., the Huber-Strassen derivative) for this model; in particular, we have

$$\pi_v(\omega) = \max\{c', (1-\epsilon)(1-r^2)(1-2r\cos(\omega)+r^2)^{-1}\}, \quad \omega \in [-\pi,\pi], \tag{38}$$

where c' is defined below (8). Straightforward analysis yields that c' is the solution to the equation

$$(1-\epsilon)D_\theta(m_0)\Big|_{\theta = c'/(1-\epsilon)} + \epsilon = c'. \tag{39}$$

The worst-case spectral density of (38) is illustrated in Fig. 1 for the case $r = .5$ and $\epsilon = .25$. Note from this illustration that, since $\frac{1}{2\pi}\int \pi_v d\lambda = 1$, it follows that c' is monotonically increasing with ϵ for $(1-r)/(1+r) \leq c' < 1$, and that c' will be identically 1 for all $\epsilon \geq \epsilon_{max}$,

where $(1-\epsilon_{max})(1+r)/(1-r) = 1$; i.e. $\epsilon_{max} \sim 2r/(1+r)$. Thus, for all

$\epsilon \geq \epsilon_{max}$, the worst-case source is memoryless (i.e., white). Note also

that, for general ϵ, a fraction

$$\frac{c'}{2\pi} \int\limits_{\{\pi_v = c'\}} \lambda\,(d\omega) \tag{40}$$

of the worst-case source power can be thought of as being due to a memory-

less component. This fraction along with the value of c' are plotted versus

ϵ in Fig. 2 for the case $r = .75$ (for which $\epsilon_{max} = 6/7$). Note that, for

$\epsilon = .1$, about 12% of the source power is due to the memoryless component and,

for $\epsilon = .3$, about 45% is due to the memoryless component.

The rate-distortion function for the worst-case source can be derived

straightforwardly from (12) and (13). After some analysis we have

$$R_q(D) = \frac{1}{2} \log\,(c'/D) + R_{m_0}(D_\theta(m_0))\Big|_{\theta\,=\,c'/(1-\epsilon)} \tag{41}$$

for $0 < D \leq c'$, and $R_q(D)$ is given parametrically by

$$D_\theta(q) = (1-\epsilon)D_\gamma(m_0)\Big|_{\gamma\,=\,\theta/(1-\epsilon)} + \epsilon \tag{42}$$

and

$$R_q(D_\theta(q)) = R_{m_0}(D_\gamma(m_0))\Big|_{\gamma\,=\,\theta/(1-\epsilon)}, \tag{43}$$

for $c' < D \leq 1$, which corresponds to $c' < \theta \leq (1-\epsilon)(1+r)/(1-r)$. Here

$D_\gamma(m_0)$ and $R_{m_0}(D_\gamma(m_0))$ are defined by (33) through (37) and q denotes the

measure corresponding to the Lebesgue density π_v (i.e., $dq = \pi_v d\lambda$). The

rate-distortion function of (41) through (43) is plotted in Fig. 3 for the case $r = .75$ and for several values of ϵ. Note that, as ϵ increases from zero to ϵ_{max}, the convexity of the worst-case curve becomes less pronounced. Of course, the $\epsilon = 0$ curve corresponds to the uncontaminated Gauss-Markov source, and the $\epsilon \geq \epsilon_{max}$ curve corresponds to a memoryless source which is universally the worst case.

It is interesting to note that (39) and (41) through (43) also apply to mixtures with nominal spectra other than the Gauss-Markov. That is, if $\int \pi_0 d\lambda = 2\pi$ where $\pi_0 = dm_0/d\lambda$, then the mixture model with nominal measure m_0 has worst-case spectral density given by (8) and (39), and the corresponding rate-distortion function is given by (41) through (43) with $D_\gamma(m_0)$ and $R_{m_0}(D_\gamma(m_0))$ being the rate-distortion equations for m_0. Thus, behavior similar to that of Fig. 3 would be expected for other mixtures. For example, for $\epsilon \geq 1 - \operatorname*{ess\,sup}_{\omega \in \Omega} \pi_0(\omega)$, such a mixture class would contain a memoryless source.

V. EXTENSION TO BANDLIMITED CONTINUOUS-PARAMETER SOURCES

Theorems 1 and 2, as proven above, apply to discrete-parameter sources. Rate-distortion functions for continuous-parameter sources, however, can also be determined from (12) and (13) provided their spectral densities are essentially bounded (see Berger [2, pp. 116-122]). Thus, if we assume that $\underline{X} = \{X_t; t \in \mathbf{R}^n\}$ is a homogeneous continuous-parameter source with spectral measure $m \in \mathcal{M}_v$, a capacity class on $(\mathbf{R}^n, \mathcal{B})$, we should be able to prove results analogous to Theorems 1 and 2. This is the case if we make the additional restriction that $v(\Omega^c) = 0$ for some compact $\Omega \subseteq \mathbf{R}^n$. This restriction is equivalent to assuming that all sources have a common (finite) bandlimit and is sufficient to assure that (12) and (13) apply to all source spectra in $\mathcal{M}_v$ and to apply the existing Huber-Strassen theory to differentiate v with respect to Lebesgue measure on Ω. We thus have

Theorem 3: Theorems 1 and 2 hold for n-continuous-parameter sources provided Ω is a compact subset of $\mathbf{R}^n$.

It should be noted that Huber-Strassen derivatives of capacities with respect to σ-finite (and not finite) measures can be constructed [24]; thus Theorem 3 can possibly be extended to nonbandlimited sources provided that the definition of π_v is appropriately modified. However, several of the most useful examples of capacity classes (e.g., the ϵ-mixtures and variation neighborhoods) fail to be capacity classes when Ω is not compact.

ACKNOWLEDGEMENT

References

1. D. J. Sakrison, "The rate distortion function for a class of sources," _Inform. and Control_, vol. 15, pp. 165-195, 1969.

2. T. Berger, _Rate Distortion Theory: A Mathematical Basis for Data Compression_. Prentice-Hall: Englewood Cliffs, New Jersey, 1971.

3. D. J. Sakrison, "The rate of a class of random processes," _IEEE Trans. Inform. Theory_, vol. IT-16, pp. 10-16, January 1970.

4. G. Choquet, "Theory of capacities," _Ann. Inst. Fourier_, vol. 5, pp. 131-292, 1953/54.

5. P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," _Ann. Statist._, vol. 1, pp. 251-263, 1973.

6. H. V. Poor, "Minimax linear smoothing for capacities," _Ann. Prob._, (to appear).

7. E. Wong, _Stochastic Processes in Information and Dynamical Systems_. McGraw-Hill: New York, 1971.

8. P. J. Huber and V. Strassen, "Correction to 'Minimax tests and the Neyman-Pearson lemma for capacities'," _Ann. Statist._, vol. 2, pp. 223-224, 1974.

9. P. J. Huber, "A robust version of the probability ratio test," _Ann. Math. Stat._, vol. 36, pp. 1753-1758, 1965.

10. R. D. Martin and S. C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," _IEEE Trans. Inform. Theory_, vol. IT-17, pp. 50-56, January 1971.

11. N. Dunford and J. T. Schwartz, _Linear Operators — Part I: General Theory_. Interscience: New York, 1967.

12. H. V. Poor, "On robust Wiener filtering," _IEEE Trans. Automatic Control_, vol. AC-25, pp. 531-536, June 1980.

13. P. J. Huber, "Robust confidence limits," _Z. Wahr. verw. Geb._, vol. 10, pp. 269-278, 1968.

14. H. Rieder, "Least favorable pairs for special capacities," _Ann. Statist._, vol. 5, pp. 909-921, 1977.

15. V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," _Telecomm. and Radio Eng._, vol. 30/31, pp. 58-64, 1976.

16. S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," _IEEE Trans. Inform. Theory_, vol. IT-27, pp.   , March 1981.

17. S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171-185, 1977.

18. K. S. Vastola and H. V. Poor, "On generalized band models in robust detection and filtering," *Proc. 14th Annual Conf. on Inform. Sciences and Systems*, Princeton University, Princeton, NJ, March 1980, pp. 1-5.

19. V. Strassen, "The existence of probability measures with given marginals," *Ann. Math. Stat.*, vol. 36, pp. 423-439, 1965.

20. J. F. Hayes, A. Habibi, P. A. Wintz, "Rate distortion function for a source model of images," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 507-509, 1970.

21. H. L. Royden, *Real Analysis*. Macmillan: Toronto, 1968.

22. S. Kullback, *Information Theory and Statistics*. Wiley: New York, 1959.

23. L. Lewin, *Dilogarithms and Associated Functions*. Macdonald: London, 1958.

24. K. S. Vastola and H. V. Poor, "On Choquet capacities and their derivatives with respect to σ-finite measures," *Abstracts of Papers: 1981 IEEE International Symposium on Information Theory*, Santa Monica, CA, February 1981, p. 91.

List of Footnotes

1. For compactness of notation we will write $\{f > \theta\}$ to denote $\{\omega \in \Omega \mid f(\omega) > \theta\}$.

2. Note that this set function (and the one defined in Example 2) is discontinuous from above at the null set ϕ. However, for $\Omega = [-\pi,\pi]^n$, this discontinuity does not violate the property that v must be continuous from above on closed sets (i.e., Property (iv)) since, in a compact separable metric space, there is no sequence of closed sets converging down to the null set (Dunford and Schwartz [11, pp. 30-31]).
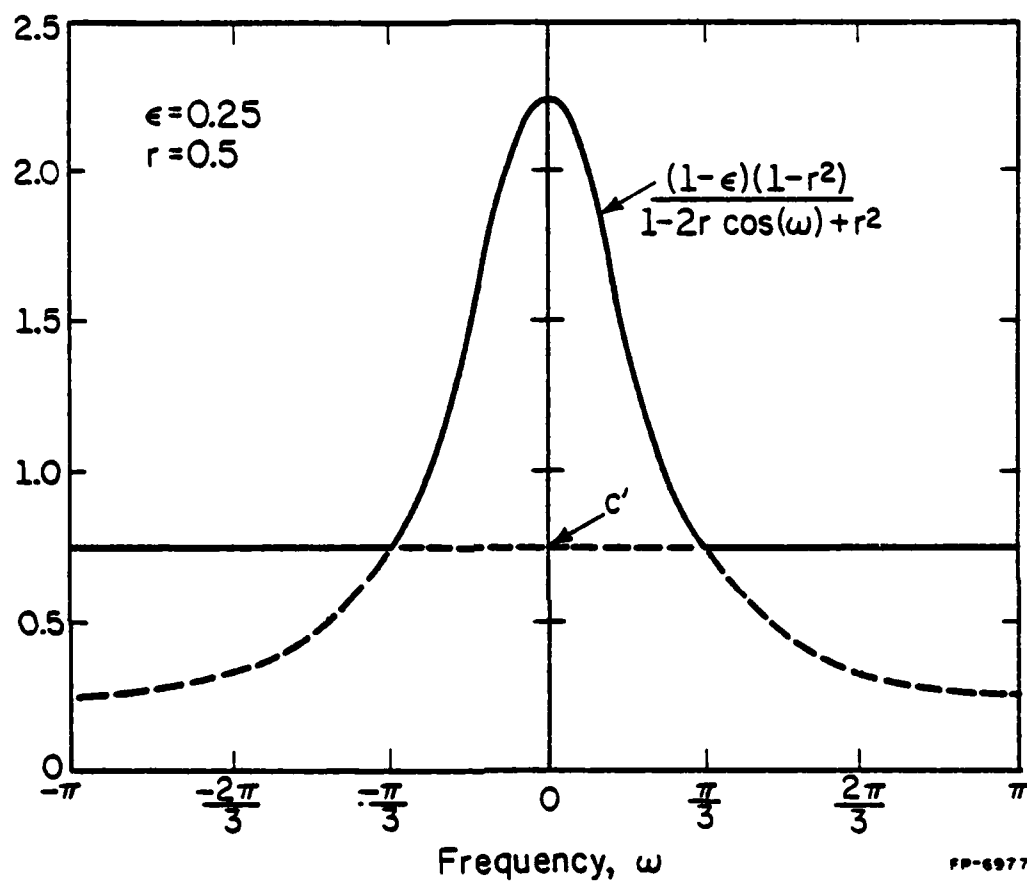
List of Figures

Fig. 1 – Worst-case spectral density for a contaminated-
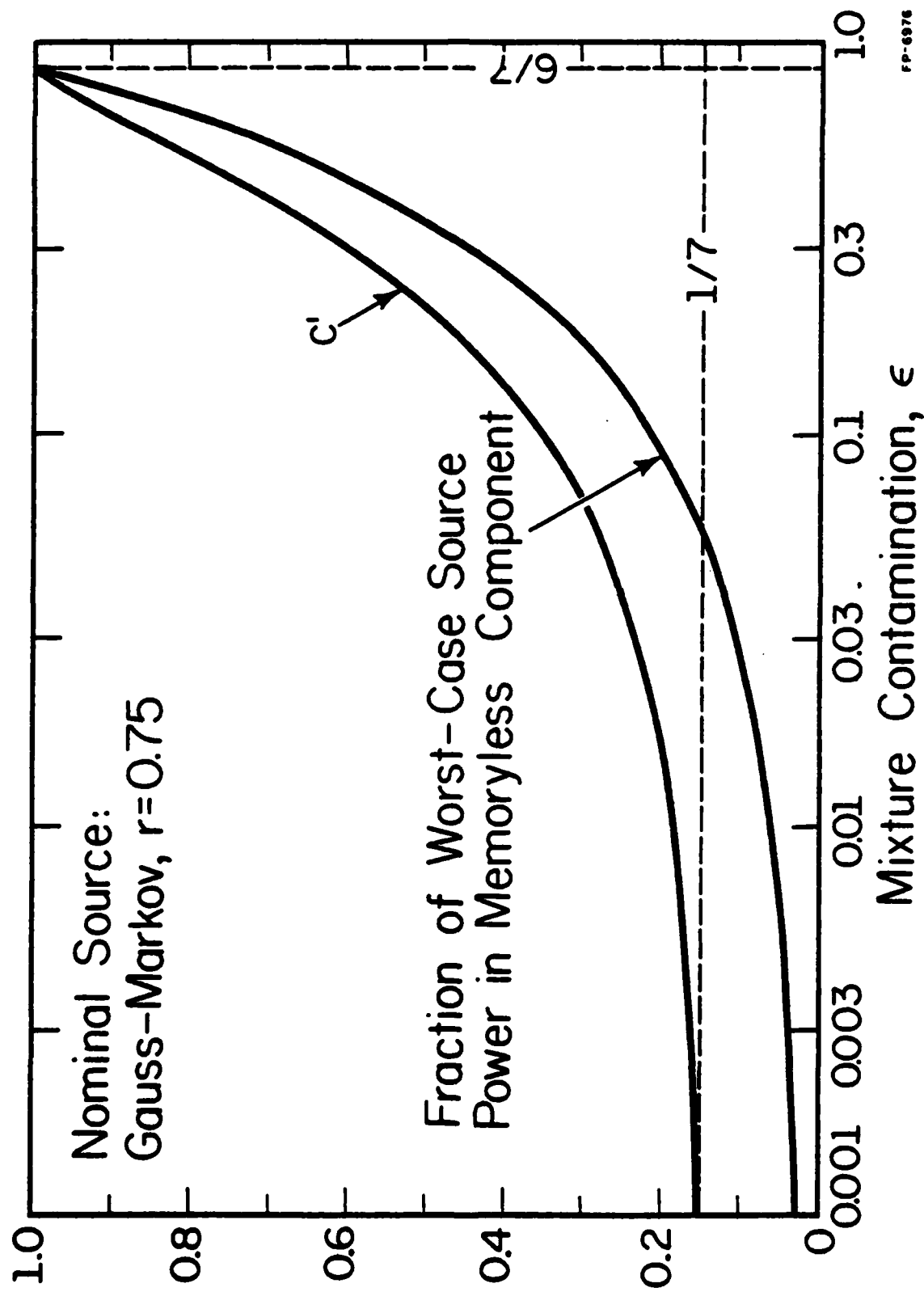mixture class ($\epsilon = .25$) with a nominal Gauss-
Markov source ($r = .5$).

Fig. 2 - Worst-case spectrum parameter c' and fraction of worst-case spectrum power contained in memoryless component versus ε; mixture-contaminated Gauss-Markov source (r = .75).
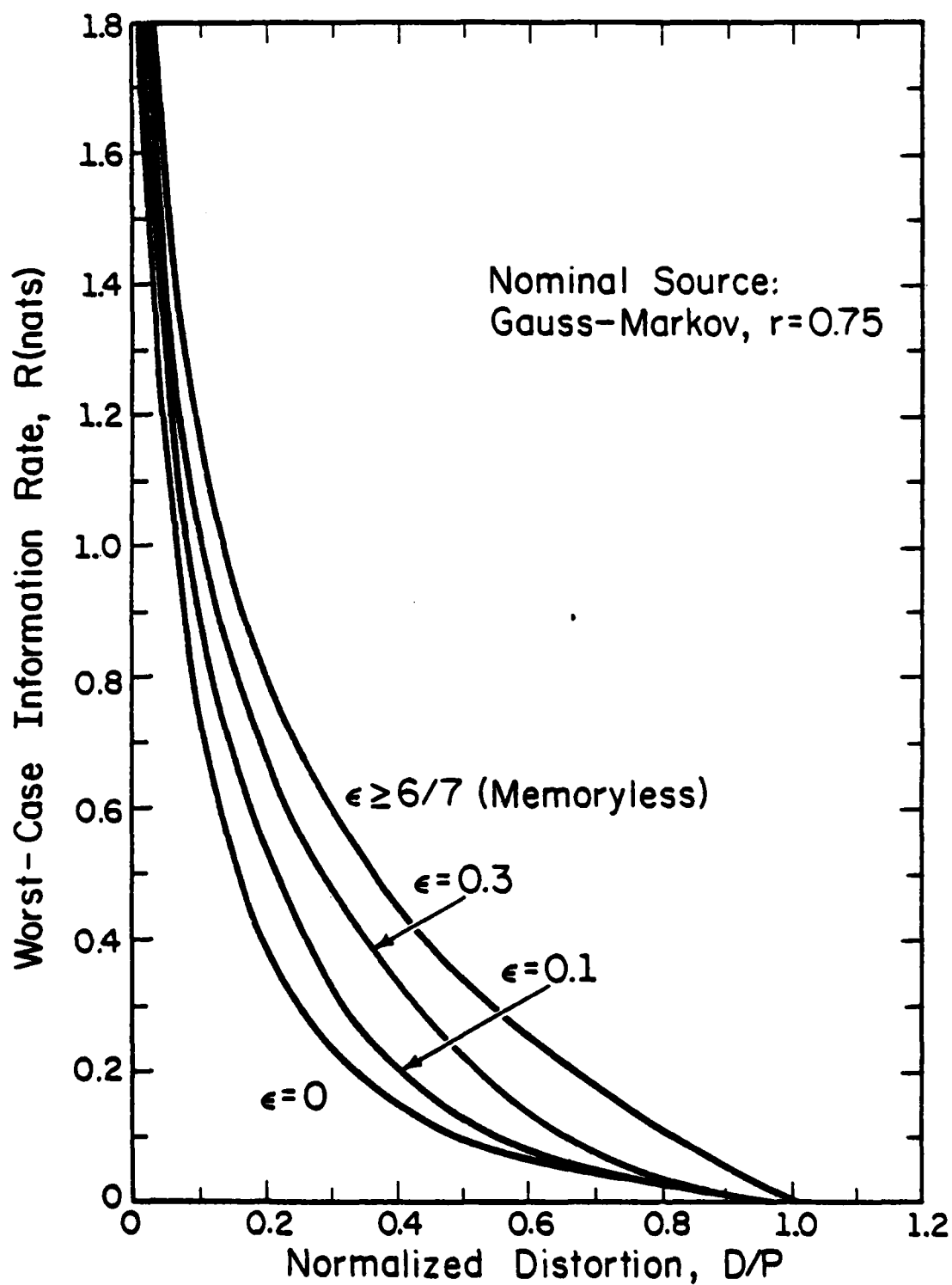
Fig. 3 - sup $R_m(D)$ for mixture-contaminated Gauss-Markov
$\quad m \in \mathcal{M}$
$\quad\quad$ source (r = .75).